

## SCIKIT LEARN – MASHINA TA'LIMI UCHUN KUTUBXONALAR TO'PLAMI

*Amirov Islombek Dilshodbek o'g'li<sup>1</sup>*

*Andijon davlat universiteti, Axborot texnologiyalari kafedrası*

*Axmadaliyev Doniyorbek Kambaraliyevich<sup>2</sup>*

*Andijon shahridagi Prezident maktabi*

*E-mail: [islombek.a.d.1606.@gmail.com](mailto:islombek.a.d.1606.@gmail.com)*

**Annotatsiya:** Scikit-learn kutubxonasi Python dasturlash tilida Sun'iy intellektning mashina ta'limi (Machine learning) ga oid muommolarini hal qilish uchun keng tarqalgan tanlovlardan biridir. Ya'ni Sun'iy intellektni mustaqil o'rganish keng ko'lamli algoritmlarni taqdim etadilar. (Machine learning) DataSetdagi belgilar mavjudligini nazarda tutmaydi, foydali ma'lumotlarni tasodifiy ma'lumlardan qanday qilib olishni o'rgatadi. Python Kutubxonaning afzaliklaridan biri u bir nechta keng tarqalgan matematik kutubxonalar asosida va ularni birlashtiradi. Scikit-learn ya'ni kompyuter ta'limini o'rganish algoritmlari tadqiqot uchun mashina ta'limini birinchi o'rganayotgan yangi boshlovchilar uchun, sanoat tizimlari va boshqa soxalar uchun ham keng qo'llaniladi.

**Kalit so'zlar:** Mashina ta'limi, Sun'iy intellekt, Python, Dasturlash, Statistika.

**Scikit-learn:** maxsus kutubxonalari:

**Numpy:** matematik operatsiyalar va tensor operatsiyalar.

**Scipy:** Ilmiy va texnik hisoblashlar.

**Sympy:** Belgilar matematikasi.

**IPython:** Python uchun interaktiv konsul.

**Matplotlib:** Ma'lumotlarni ko'rish.

**Pandas:** Ma'lumotlarni qayta ishlash, va tahlil qilish.

## I. Kirish

Mashina ta'limi, kompyuterni huddi insonlar kabi avvalgi tajribalardan hulosa (algoritm moduli yaratib) chiqargan holda masalalarga yechim taklif qilish yoki mustaqil qaror chiqarishga imkon beruvchi Suni'y intellekt ilovalaridan biri hisoblanadi. Mashina ta'limi tushunchasi birinchi marta 1860-yillarda ishlatila boshlangan, 20 yil avval esa u har bir sohada – fanlar aro – ishlatila boshladi. Unda ehtimollar nazariyasi, statistika va boshqalar mavjud. Kompyuterning texnikasi yangilanishi bilan hisoblash tezligi o'sishda davom etmoqda. Yaqin kelajakda bu bizning hayotimizda muhim ahamiyat kasb eta boshlaydi.

Mashina ta'limini o'rganish jarayoni haqida quyidagilarni aytib o'tishimiz mumkin: yetakchi kompaniyalardan biri mashinani o'rganish asosida ular Google Now, Google fotosuratlarini va yana boshqalar. Masalan moliyaviy kompaniyalar banklar undan valyuta kursini bashorat qilishda foydalanishlari mumkin, ko'chmas mulk kompaniyalari esa uy narxini va boshqa buyumlarning sotilish trendi yoki ularga bo'ladigan talablarni oldindan bashorat (прогноз) qilishadi. Biz kompyuterlarga ma'lumotlarni va ma'lumotlarga mos keladigan yorliqlarni taqdim etamiz. Masalan, kompyuterga smartfon bilan oddiy bir televizorni pultini rasmini ko'rsatib qaysi biri smartfon yoki oddiy televizorni pulti ekanligini aniqlash uchun ularni farqlashni o'rganish kerak bo'ladi.

Rasmlarni ko'rsatib kompyuterga bu smartfon yoki pult ekanligini ajratib olishiga ruxsat berganimizdan so'ng ya'ni qaysi rasmlarni bitta yorliq bilan ifodalash mumkinligini bilib olsin. Ya'ni ushbu usul “nazorat ostida o'rganish” deb nomlanadi. Valyutalar kursini va uylar savdosini bashorat (прогноз) qilishda va barcha ehtimollari nazorat ostida o'qitish yordamida amalga oshiriladi, neyron to'rlari nazorat qilinadigan o'qitishning turlaridan biridir. Kompyuterga har xil yana boshqacharoq smartfon bilan pultni rasmlarini ko'rsatadigan bo'lsak lekin qaysi biri smartfon yoki pult ekanligini oldindan aytmaymiz, endi kompyuter shu ikki xil turdagi rasmlarni o'zi tahlil qilib farqlaydi, bu esa nazoratsiz (unsupervised)

o'rganish deb ataladi, va ma'lumotlarni kuzatish orqali kompyuter ushbu funksiyalarni saqlaydi, shu qoidalar nazoratsiz o'rganiladigan Mashina ta'limi. Ushbu yondashuv nazorat ostida o'qitish (supervised)larning va nazoratsiz o'rganishni birlashtiradi va bu “yarim nazorat ostidagi o'rganish” deb ham nomlanadi. Kompyuterni harakatlarini o'rganish nuqtai nazardan “mustahkamlashni o'rganish” deb nomlangan mashinada o'qitish usullari mavjud.

Masalan: Qo'lyozmadagi raqamlarni tanib olish.

Ushbu misol scikit-learn-dan qo'lda yozilgan raqamlarning tasvirlarini tanib olish uchun qanday ishlatilishini ko'rsatadi.

(Algoritm 1) ning 8 va 9 qatorlarida raqamlarni aniqlab olish uchun *Sklearn* ning *datasets*, *svm* va *metrics* lari import qilib olinmoqda

### Algoritm 1.

```
print(__doc__)

# Author: Gael Varoquaux <gael dot varoquaux at normalesup dot org>
# License: BSD 3 clause

# Standard scientific Python imports
import matplotlib.pyplot as plt

# Import datasets, classifiers and performance metrics
from sklearn import datasets, svm, metrics
from sklearn.model_selection import train_test_split
```

Python interaktiv muhiti uchun avtomatik ravishda yaratilgan modullar

### Raqamli ma'lumotlar to'plami.

Ma'lumotlar to'plami raqamlarning 8x8 pikselli rasmlaridan iborat. Ma'lumotlar to'plamining tasvirlar atributi har bir rasm uchun 8x8 qatorli kulranglarni saqlaydi. Dastlabki 4 ta rasmni tasavvur qilish uchun ushbu massivlardan foydalanamiz.

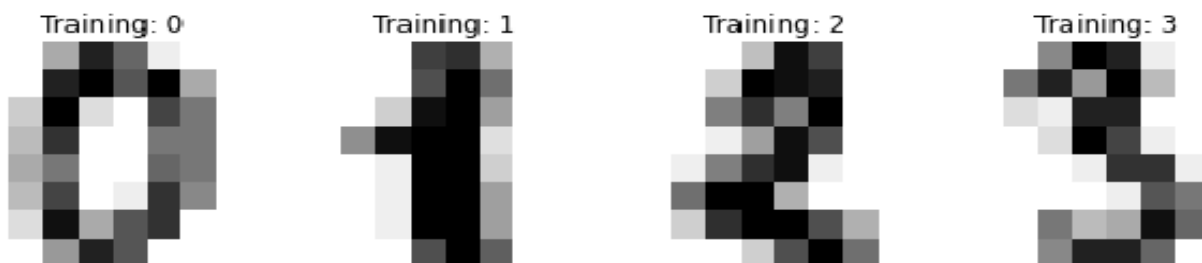
Ma'lumotlar to'plamining maqsadli atributi har bir rasm ko'rsatadigan raqamni saqlaydi va bu quyidagi 4 ta uchastkaning sarlavhasiga kiritilgan.

**Izoh:** agar biz rasm fayllaridan ishlayotgan bo'lsak (masalan, 'png'files), ularni quyidagicha yuklar edik: `func: matplotlib.pyplot.imshow`.

## Algoritm 2.

```
digits = datasets.load_digits()

_, axes = plt.subplots(nrows=1, ncols=4, figsize=(10, 3))
for ax, image, label in zip(axes, digits.images, digits.target):
    ax.set_axis_off()
    ax.imshow(image, cmap=plt.cm.gray_r, interpolation='nearest')
    ax.set_title('Training: %i' % label)
```



```
digits.target

array([0, 1, 2, ..., 8, 9, 8])
```

## Tasnif

Ushbu ma'lumotlarga tasniflagichni qo'llash uchun biz (8, 8) shakldan (8, 8) kulrangdagi har ikki o'lchovli massiv (64,) shaklga aylantirib, rasmlarni to'plashimiz kerak. Keyinchalik, barcha ma'lumotlar to'plami shaklga ega bo'ladi (`n_samples`, `n_features`).

**Bu erda:** `n_samples` - rasmlar soni va `n_feature` - har bir rasmdagi piksellarning umumiy soni.

Keyin biz ma'lumotlarni mashq va sinov pastki qismlariga ajratishimiz va mashq namunalariga qo'llab-quvvatlash vektor tasniflagichini joylashtirishimiz mumkin. O'rnatilgan tasniflagich keyinchalik sinov pastki qismidagi namunalar uchun raqamning qiymatini taxmin qilish uchun ishlatilishi mumkin.

### Algorithm 3.

```
# flatten the images
n_samples = len(digits.images)
data = digits.images.reshape((n_samples, -1))

# Create a classifier: a support vector classifier
clf = svm.SVC(
    gamma=0.001)

# Split data into 50% train and 50% test subsets
X_train, X_test, y_train, y_test = train_test_split(
    data, digits.target, test_size=0.5, shuffle=False)

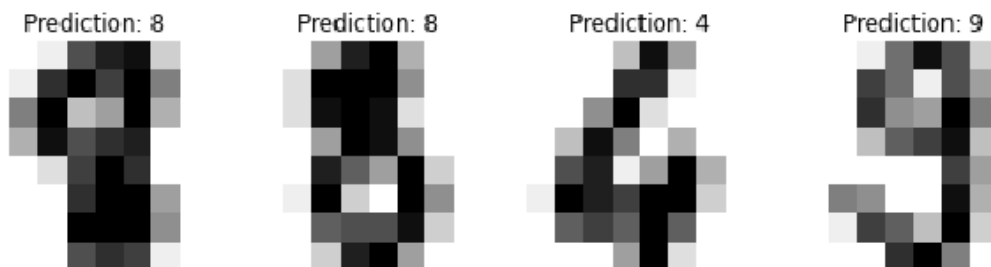
# Learn the digits on the train subset
clf.fit(X_train, y_train)

# Predict the value of the digit on the test subset
predicted = clf.predict(X_test)
```

Quyida biz dastlabki 4 ta namunani tasavvur qilamiz va ularning taxmin qilingan raqam qiymatini sarlavhada ko'rsatamiz.

### Algorithm 4.

```
_, axes = plt.subplots(nrows=1, ncols=4, figsize=(10, 3))
for ax, image, prediction in zip(axes, X_test, predicted):
    ax.set_axis_off()
    image = image.reshape(8, 8)
    ax.imshow(image, cmap=plt.cm.gray_r, interpolation='nearest')
    ax.set_title(f'Prediction: {prediction}')
```



: func: ~ sklearn.metrics.classification\_report asosiy tasnif o'lchovlarini ko'rsatadigan matnli hisobot tuzadi. Ushbu ko'rinish sonlarni foiz ko'rinishida aniqlab, chiqarib beradi.

## II. Algoritm

5.

```
print(f"Classification report for classifier {clf}:\n"
      f"{metrics.classification_report(y_test, predicted)}\n")
```

Classification report for classifier SVC(gamma=0.001):

	precision	recall	f1-score	support
0	1.00	0.99	0.99	88
1	0.99	0.97	0.98	91
2	0.99	0.99	0.99	86
3	0.98	0.87	0.92	91
4	0.99	0.96	0.97	92
5	0.95	0.97	0.96	91
6	0.99	0.99	0.99	91
7	0.96	0.99	0.97	89
8	0.94	1.00	0.97	88
9	0.93	0.98	0.95	92
accuracy			0.97	899
macro avg	0.97	0.97	0.97	899
weighted avg	0.97	0.97	0.97	899

Bundan tashqari, biz chalkashlik matritsasini tuzishimiz mumkin <confusion\_matrix>. haqiqiy raqam qiymatlari va bashorat qilingan raqam qiymatlari.

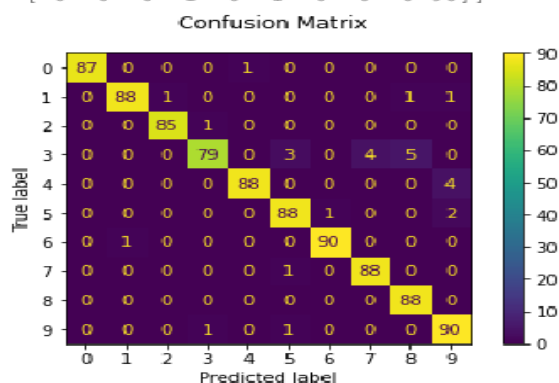
## Algoritm 6.

```
disp = metrics.plot_confusion_matrix(clf, X_test, y_test)
disp.figure_.suptitle("Confusion Matrix")
print(f"Confusion matrix:\n{disp.confusion_matrix}")

plt.show()
```

Confusion matrix:

```
[[ 87  0  0  0  1  0  0  0  0  0]
 [ 0 88  1  0  0  0  0  0  1  1]
 [ 0  0 85  1  0  0  0  0  0  0]
 [ 0  0  0 79  0  3  0  4  5  0]
 [ 0  0  0  0 88  0  0  0  0  4]
 [ 0  0  0  0  0 88  1  0  0  2]
 [ 0  1  0  0  0  0 90  0  0  0]
 [ 0  0  0  0  0  1  0 88  0  0]
 [ 0  0  0  0  0  0  0  0 88  0]
 [ 0  0  0  1  0  1  0  0  0 90]]
```



Rasm 3. Sonlar foiz ko'rinishda aks ettirilgan

### III. NATIJALAR

Ushbu maqolada biz muammolarni hal qilish uchun mashina ta'lim texnikasi scikit-learnni Python yordamida ma'lumotlarga qanday qo'llanilishi mumkinligini oddiy misollar bilan tasvirladik. Mashinani o'rganish vazifalari uchun kuzatuvchi Python kodi sodda hisoblanib, muammoli ma'lumotlarga to'g'ri tayyorgarlik ko'rish, muammo uchun to'g'ri modelni tanlash va natijalarni talqin qilish uchun qulaydir. Scikit-learn Python uchun bepul va ochiq kodli kompyuterni o'rganish kutubxonasi. Ushbu kutubxona ma'lumotlarni olish va ma'lumotlarni tahlil qilish uchun qulay foydalanish vositalarini taqdim etadi. Asosan Scikit-learn nazorat va nazorat qilinmagan muammolar uchun ilg'or mashina ta'lim algoritmlarni katta sonini beradi.

### IV. Xulosa

Scikit-learn nazorat va har ikkala mashina ta'lim algoritmlarni turli xil kutubxonalar yordamida, vazifaga yo'naltirilgan interfeyslardan biri shunday qilib oson va taqqoslash imkoni berilgan. Qo'llash uchun usullari ilmiy Python muhitiga tayangani uchun, uni osonlik bilan statistik ma'lumotlarni tahlil qilish. Eng muhimi, yuqori darajadagi tilda amalga oshirilayotgan algoritmlardan foydalanish mumkin. Hozirgi paytda scikite learn bizning hayotimizga kirib kela boshlamoqda asta sekinlik bilan hozirgi kunda ham keying yillarda ham ilmiy ishlar olib boriladi va insoniyatga bu sohada yanada yangilanishlar g'oyalar kela boshlaydi. Biz hozirgi yozgan kodimizda har xil yozuvdagi shakli raqamlardi tahminiy yo'l bilan ajratib olish mumkinligini bilib beruvchi dasturni ishga tushirdik. Bundan yana boshqa har xil turdagi loyixa tuzib kodlarni yozib dastur tuzishimiz mumkin.

### V. Foydalanilgan Adabiyotlar:

1. "Scikite-learn" internet kanali.
2. Python kutubxonasi qo'llanmassidan.
3. <https://www.oreilly.com>
4. <https://medium.com>